

Blind Beam-Hardening Correction from Poisson Measurements

Renliang Gu^{a)} and Aleksandar Dogandžić^{b)}

Iowa State University, Center for Nondestructive Evaluation, 1915 Scholl Road, Ames, IA 50011, USA

^{a)}renliang@iastate.edu ^{b)}corresponding author: ald@iastate.edu

Abstract. We develop a sparse image reconstruction method for Poisson-distributed polychromatic X-ray computed tomography (CT) measurements under the blind scenario where the material of the inspected object and the incident energy spectrum are unknown. We employ our mass-attenuation spectrum parameterization of the noiseless measurements and express the mass-attenuation spectrum as a linear combination of B-spline basis functions of order one. A block coordinate-descent algorithm is developed for constrained minimization of a penalized Poisson negative log-likelihood (NLL) cost function, where constraints and penalty terms ensure nonnegativity of the spline coefficients and nonnegativity and sparsity of the density map image; the image sparsity is imposed using a convex total-variation (TV) norm penalty term. This algorithm alternates between a Nesterov's proximal-gradient (NPG) step for estimating the density map image and a limited-memory Broyden-Fletcher-Goldfarb-Shanno with box constraints (L-BFGS-B) step for estimating the incident-spectrum parameters. To accelerate convergence of the density-map NPG steps, we apply function restart and a step-size selection scheme that accounts for varying local Lipschitz constants of the Poisson NLL. Real X-ray CT reconstruction examples demonstrate the performance of the proposed scheme.

BACKGROUND AND INTRODUCTION

X-ray photons arriving at a detector can be well modeled using a Poisson process [1, 2], where the energy deposited by each photon follows a distribution that is proportional to the incident energy spectrum. Hence, the measurements collected by photon-counting and energy-integrating detectors are modeled using Poisson and compound-Poisson distributions, respectively [3, 4]. However, the compound-Poisson distribution is complex and does not have a closed-form expression, which motivates introduction of approximations such as Poisson [5] and lognormal [6–8]. Regardless of the statistical measurement models, most polychromatic X-ray computed tomography (CT) reconstruction methods assume known X-ray spectrum and materials (i.e., known mass-attenuation function), with goal to maximize the underlying likelihood function or its regularized version [5, 9]. However, the X-ray spectrum measurements based on the semiconductor detectors are usually distorted by charge trapping, escape events, and other effects [10] and the corresponding correction requires highly collimated beam and special procedures [11, 12]. Knowing the mass-attenuation function can be challenging as well when the inspected material is unknown, or the inspected object is made of compound or mixture with unknown percentage of each constituent.

Van Gompel et al. [6] consider a “blind” scenario for lognormal measurement model with *unknown* incident spectrum and materials and employ the K -means clustering method to initially associate pixels to the materials and then alternate between material segmentation and updating the relative density map, incident X-ray spectrum, and mass attenuation coefficients for each material. The methods in [6] are based on the standard photon-energy parameterization, employ an excessive number of parameters, and suffer from numerical instability [13]. Indeed, iteratively updating excessive numbers of non-identifiable parameters *does not* lead to robust reconstruction schemes.

In this paper, we adopt our parsimonious mass-attenuation spectrum parameterization [8] and develop a blind sparse density-map reconstruction scheme from measurements corrupted by Poisson noise. The Poisson noise model is a good approximation for the more precise compound Poisson distribution for energy measurements [7, 14]. Note that [8] considers the lognormal noise model (rather than Poisson) and employs different sparsifying transform and reconstruction scheme than the approach presented here.

We introduce the notation: I_N , $\mathbf{1}_{N \times 1}$, and $\mathbf{0}_{N \times 1}$ are the identity matrix of size N and the $N \times 1$ vectors of ones and zeros (replaced by I , $\mathbf{1}$, and $\mathbf{0}$ when the dimensions can be inferred easily); $|\cdot|$, $\|\cdot\|_p$, and “ T ” are the absolute value, ℓ_p norm, and transpose, respectively. Furthermore, “ \succeq ” is the elementwise version of “ \geq ”, $\lceil x \rceil$ is the smallest integer larger than or equal

to $x \in \mathbb{R}$, $\mathbb{I}_{[0, +\infty)}(\boldsymbol{\alpha}) = \begin{cases} 0, & \boldsymbol{\alpha} \succeq \mathbf{0} \\ +\infty, & \text{otherwise} \end{cases}$ is the nonnegativity indicator function for a real-valued vector $\boldsymbol{\alpha}$, and $\mathbf{a}^L(s)$ is

the Laplace transform of a vector function $\mathbf{a}(\kappa)$: $\mathbf{a}^L(s) = \int \mathbf{a}(\kappa) e^{-s\kappa} d\kappa$. Define also the set of nonnegative real numbers as $\mathbb{R}_+ = [0, +\infty)$, the elementwise logarithm $\ln_{\circ}(s) = [\ln s_1, \dots, \ln s_N]^T$, and Laplace transform $\mathbf{a}_{\circ}^L(s) = (\mathbf{a}^L(s_n))_{n=1}^N$ obtained by stacking $\mathbf{a}^L(s_n)$ columnwise, where $s = [s_1, s_2, \dots, s_N]^T$.

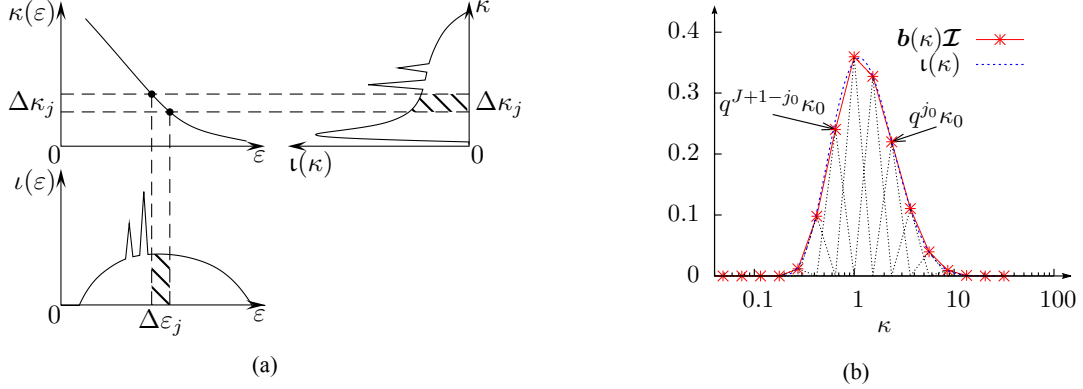


FIGURE 1. (a) Mass-attenuation spectrum $\iota(\kappa)$ obtained by combining the mass attenuation $\kappa(\varepsilon)$ and incident spectrum $\iota(\varepsilon)$, and (b) its B1-spline expansion.

In the following, we review our polychromatic X-ray CT model and describe the shift ambiguity of the mass-attenuation spectrum.

Polychromatic X-ray CT Model Using Mass-Attenuation Spectrum

By the *exponential law of absorption* [15], the fraction $d\mathcal{I}/\mathcal{I}$ of plane wave intensity lost in traversing an infinitesimal thickness $d\ell$ at Cartesian coordinates (x, y) is proportional to $d\ell$:

$$\frac{d\mathcal{I}}{\mathcal{I}} = -\mu(x, y, \varepsilon) d\ell \quad (1a)$$

where $\mu(x, y, \varepsilon)$ is the attenuation. To obtain the intensity decrease along a straight-line path $\ell = \ell(x, y)$ at photon energy ε , integrate (1a) along ℓ : $\mathcal{I}^{\text{out}}(\varepsilon) = \mathcal{I}^{\text{in}}(\varepsilon) \exp \left[-\int_{\ell} \mu(x, y, \varepsilon) d\ell \right]$, where $\mathcal{I}^{\text{out}}(\varepsilon)$ and $\mathcal{I}^{\text{in}}(\varepsilon)$ are the emergent and incident X-ray signal energies, respectively, at photon energy ε . We model the attenuation $\mu(x, y, \varepsilon)$ of the inspected object consisting of a single material using the following *separable form* [16, Sec. 6]:

$$\mu(x, y, \varepsilon) = \kappa(\varepsilon)\alpha(x, y). \quad (1b)$$

Here, $\kappa(\varepsilon) > 0$ is the *mass-attenuation coefficient* of the material, a function of the photon energy ε (illustrated in Fig. 1a) and $\alpha(x, y) \geq 0$ is the density map of the object. For a monochromatic source at photon energy ε , $\ln[\mathcal{I}^{\text{in}}(\varepsilon)/\mathcal{I}^{\text{out}}(\varepsilon)]$ is a linear function of $\alpha(x, y)$, which is a basis for traditional linear reconstruction. However, X-rays generated by vacuum tubes are not monochromatic [17], and we cannot transform the underlying noiseless measurements to a linear model *unless* we know perfectly the incident energy spectrum $\iota(\varepsilon)$ and mass attenuation of the inspected material $\kappa(\varepsilon)$.

Since the mass attenuation $\kappa(\varepsilon)$ and incident spectrum density $\iota(\varepsilon)$ are both functions of ε (see Fig. 1a), we combine the variations of these two functions and write the incident intensity \mathcal{I}^{in} and the noiseless measurement collected by an energy integral detector upon traversing a straight line $\ell = \ell(x, y)$ as [8]

$$\mathcal{I}^{\text{in}} = \iota^{\text{L}}(0), \quad \mathcal{I}^{\text{out}} = \iota^{\text{L}} \left(\int_{\ell} \alpha(x, y) d\ell \right) \quad (2)$$

where $\iota^{\text{L}}(s) = \int \iota(\kappa) e^{-s\kappa} d\kappa$ is the Laplace transform of the *mass-attenuation spectrum* $\iota(\kappa)$, which represents the density of the incident X-ray energy at attenuation κ ; here, $s > 0$, in contrast with the traditional Laplace transform where s is generally complex. For invertible $\kappa(\varepsilon)$ with differentiable inverse function $\varepsilon(\kappa)$, $\iota(\kappa) \triangleq \iota(\varepsilon(\kappa)) |\varepsilon'(\kappa)| \geq 0$, with $\varepsilon'(\kappa) = d\varepsilon(\kappa)/d\kappa$. All $\kappa(\varepsilon)$ encountered in practice can be divided into $M + 1$ piecewise-continuous segments $\{(e_m, e_{m+1})\}_{m=0}^M$, where $\kappa(\varepsilon)$ in each segment is a differentiable monotonically decreasing function of ε . In this case, (2) holds and the expression for $\iota(\kappa)$ can be easily extended to this scenario, yielding

$$\iota(\kappa) \triangleq \sum_{m=0}^M 1_{(u_m, v_m)}(\kappa) \iota(\varepsilon_m(\kappa)) |\varepsilon'_m(\kappa)| \quad (3)$$

where $1_{(u_m, v_m)}(\kappa)$ is an indicator function that takes value 1 when $\kappa \in (u_m, v_m)$ and 0 otherwise. The range and inverse of $\kappa(\varepsilon)$ within (e_m, e_{m+1}) are (u_m, v_m) and $\varepsilon_m(\kappa)$, respectively, with $u_m \triangleq \inf_{\varepsilon \nearrow e_{m+1}} \kappa(\varepsilon) < v_m \triangleq \sup_{\varepsilon \searrow e_m} \kappa(\varepsilon)$.

Observe that the mass-attenuation spectrum $\iota(\kappa)$ is nonnegative for all κ :

$$\iota(\kappa) \geq 0. \quad (4)$$

Due to its nonnegative support and range, $\iota^L(s)$ is a decreasing function of s . The function $(\iota^L)^{-1}$ converts the noiseless measurement \mathcal{I}^{out} in (2), which is a nonlinear function of the density map $\alpha(x, y)$, into a noiseless linear “measurement” $\int_{\ell} \alpha(x, y) d\ell$. The $(\iota^L)^{-1} \circ \exp(-\cdot)$ mapping corresponds to the *linearization function* in [18] and converts $-\ln \mathcal{I}^{\text{out}}$ into a noiseless linear “measurement” $\int_{\ell} \alpha(x, y) d\ell$.

Approximate $\iota(\kappa)$ with a linear combination of J B-spline basis functions of order one (termed B1 splines, illustrated in Fig. 1b) $\mathbf{b}(\kappa) \triangleq [b_1(\kappa), b_2(\kappa), \dots, b_J(\kappa)]$ [8]:

$$\iota(\kappa) = \mathbf{b}(\kappa) \mathcal{I} \quad (5a)$$

where $\mathcal{I} = [\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_J]^T \geq \mathbf{0}$ is an *unknown* $J \times 1$ vector of corresponding basis-function coefficients. In this case, the decomposition (5a) yields nonnegative elements of the spline coefficients \mathcal{I} [based on (4)] and thus allows us to impose the physically meaningful nonnegativity constraint when estimating \mathcal{I} . The spline knots are selected from a growing geometric series $\{\kappa_j\}_{j=0}^{J+1}$ with $\kappa_0 > 0$,

$$\kappa_j = q^j \kappa_0 \quad (5b)$$

and common ratio $q > 1$, which yields the B1-spline basis functions:

$$b_j(\kappa) = \begin{cases} \frac{\kappa - \kappa_{j-1}}{\kappa_j - \kappa_{j-1}}, & \kappa_{j-1} \leq \kappa < \kappa_j \\ \frac{-\kappa + \kappa_{j+1}}{\kappa_{j+1} - \kappa_j}, & \kappa_j \leq \kappa < \kappa_{j+1} \\ 0, & \text{otherwise} \end{cases} \quad \text{satisfying} \quad b_j(\kappa) = b_{j+1}(q\kappa) \quad (5c)$$

see also Fig. 1b. The geometric-series knots have a wide span from κ_0 to κ_{J+1} and compensate larger κ [i.e., higher attenuation κ_j implies an exponentially smaller $\exp(-\phi^T \alpha \kappa_j)$ term] with a geometrically wider integral range, that results in an effective approximation of (2). The common ratio q determines the resolution of the B1-spline approximation. In summary, the following three tuning constants:

$$(q, \kappa_0, J) \quad (5d)$$

define our B1-spline basis functions $\mathbf{b}(\kappa)$.

Spatial-domain discretization into p pixels corresponds to the following approximation:

$$\int_{\ell} \alpha(x, y) d\ell \approx \phi^T \alpha \quad (6)$$

where $\alpha \geq \mathbf{0}$ is an *unknown* $p \times 1$ vector representing the 2D image that we wish to reconstruct [i.e., discretized $\alpha(x, y)$] and $\phi \geq \mathbf{0}$ is a $p \times 1$ vector of weights quantifying how much each element of α contributes to the X-ray attenuation on the straight-line path ℓ . An X-ray CT scan consists of hundreds of projections with the beam intensity measured by thousands of detectors for each projection. Denote by N the total number of measurements from all projections collected at the detector array. For the n th measurement, define its discretized line integral as $\phi_n^T \alpha$; stacking all N such integrals into a vector yields $\Phi \alpha$, where $\Phi = [\phi_1 \phi_2 \dots \phi_N]^T \in \mathbb{R}^{N \times p}$ is the *projection matrix*, also called Radon transform matrix in a parallel-beam X-ray tomographic imaging system. We call the corresponding transformation, $\Phi \alpha$, the monochromatic projection of α .

Substituting (5a) and (6) into (2) for each of the N measurements yields the following expressions for the incident energy and the $N \times 1$ vector of noiseless measurements:

$$\mathcal{I}^{\text{in}}(\mathcal{I}) = \mathbf{b}^L(0) \mathcal{I}, \quad \mathcal{I}^{\text{out}}(\alpha, \mathcal{I}) = \mathbf{b}_o^L(\Phi \alpha) \mathcal{I} \quad (7)$$

where $\mathbf{b}_o^L(s) \triangleq (\mathbf{b}^L(s_n))_{n=1}^N$ is an *output basis-function matrix* obtained by stacking the $1 \times J$ vectors $\mathbf{b}^L(s_n)$ columnwise and $s = \Phi \alpha$ is the monochromatic projection.

Shift Ambiguity of the Mass-Attenuation Spectrum

By noting the κ -scaling property of the Laplace transform, $b_j(q\kappa) \xrightarrow{\mathcal{L}} \frac{1}{q} b_j^L(s/q)$ for $q > 0$, we conclude that selecting q times narrower basis functions $[b_0(\kappa), b_1(\kappa), \dots, b_{J-1}(\kappa)]$ than those in $\mathbf{b}(\kappa)$ and q times larger density map and spectral parameters ($q\alpha$ and $q\mathcal{I}$) yields the same mean output photon energy. Consequently,

$$\mathcal{I}^{\text{out}}(\alpha, [0, \mathcal{I}_2, \dots, \mathcal{I}_J]^T) = \mathcal{I}^{\text{out}}(q\alpha, q[\mathcal{I}_2, \dots, \mathcal{I}_J, 0]^T). \quad (8)$$

Hence, when \mathcal{I} has a leading zero, the noiseless signal output *remains the same* if we shift the elements of \mathcal{I} (and correspondingly the mass-attenuation spectrum) to the left, followed by scaling the new \mathcal{I} and α by q . Equivalently, when \mathcal{I} has a trailing zero, the noiseless signal output remains the same if we shift the elements of \mathcal{I} (and correspondingly the spectrum) to the right and scale the new \mathcal{I} and α by $1/q$. We refer to this property as the *shift ambiguity* of the mass-attenuation spectrum, which allows us to rearrange leading or trailing zeros in the mass-attenuation coefficient vector \mathcal{I} and position the central nonzero part of \mathcal{I} .

In the following section, we introduce the Poisson noise model and corresponding penalized negative log-likelihood (NLL) objective function and present a minimization algorithm that aims at minimizing this objective function.

PARAMETER ESTIMATION

Poisson Measurement Model

For an $N \times 1$ vector \mathcal{E} of independent Poisson measurements, the NLL in the form of generalized Kullback-Leibler divergence [19] is

$$\mathcal{L}(\alpha, \mathcal{I}) = \mathbf{1}^T [\mathcal{I}^{\text{out}}(\alpha, \mathcal{I}) - \mathcal{E}] - \mathcal{E}^T \left\{ \ln_{\circ} [\mathcal{I}^{\text{out}}(\alpha, \mathcal{I})] - \ln_{\circ} \mathcal{E} \right\} \quad (9)$$

see also (7). We now express (9) as a function of α with \mathcal{I} fixed and vice versa, which will be used to describe our estimation algorithm.

NLL of α . Recall (5a) and write the NLL of α for fixed \mathcal{I} as

$$\mathcal{L}_{\mathfrak{t}}(\alpha) = \mathbf{1}^T [\mathfrak{t}_{\circ}^{\mathfrak{L}}(\Phi\alpha) - \mathcal{E}] - \mathcal{E}^T \left\{ \ln_{\circ} [\mathfrak{t}_{\circ}^{\mathfrak{L}}(\Phi\alpha)] - \ln_{\circ} \mathcal{E} \right\} \quad (10a)$$

which corresponds to the Poisson generalized linear model (GLM) with design matrix Φ and link function equal to the inverse of $\mathfrak{t}^{\mathfrak{L}}(\cdot)$. See [20] for introduction to GLMs.

NLL of \mathcal{I} . Fix α and define $A = \mathfrak{b}_{\circ}^{\mathfrak{L}}(\Phi\alpha)$. The NLL of \mathcal{I} for fixed α reduces to

$$\mathcal{L}_A(\mathcal{I}) = \mathbf{1}^T (A\mathcal{I} - \mathcal{E}) - \mathcal{E}^T [\ln_{\circ}(A\mathcal{I}) - \ln_{\circ} \mathcal{E}] \quad (10b)$$

which corresponds to the Poisson GLM with design matrix A and identity link.

Biconvexity of $\mathcal{L}(\alpha, \mathcal{I})$. We have shown that the NLL (9) is biconvex with respect to α and \mathcal{I} in the following set:

$$\left\{ (\alpha, \mathcal{I}) \left| \mathcal{I}^{\text{out}}(\alpha, \mathcal{I}) \geq \frac{(q^{j_0} - 1)^2}{q^{2j_0} + 1} \mathcal{E}, \mathcal{I} \in \mathcal{A}, \alpha \in \mathbb{R}_+^p \right. \right\} \quad (11a)$$

which bounds $\mathcal{I}_n^{\text{out}}(\alpha, \mathcal{I})/\mathcal{E}_n$ from below for all n . Here, we constrain the region of allowed \mathcal{I} to

$$\mathcal{A} = \left\{ \mathcal{I} \in \mathbb{R}_+^J \left| \mathcal{I}_1 \leq \mathcal{I}_2 \leq \dots \leq \mathcal{I}_{J+1-j_0}, \mathcal{I}_{j_0} \geq \dots \geq \mathcal{I}_{J-1} \geq \mathcal{I}_J, \text{ and } \mathcal{I}_j \geq \mathcal{I}_{J+1-j_0}, \forall j \in [J+1-j_0, j_0] \right. \right\} \quad (11b)$$

which enforces monotonicity of the mass-attenuation spectrum $\mathfrak{t}(\kappa)$ in low- and high- κ regions (determined by a constant $j_0 \geq \lceil (J+1)/2 \rceil$) and also assume that the mid- κ region has higher spectrum than the low- κ region.

Penalized NLL Objective Function

Our goal is to compute penalized maximum-likelihood estimates of the density-map and mass-attenuation spectrum parameters (α, \mathcal{I}) by solving the following constrained minimization problem:

$$\min_{\alpha, \mathcal{I} \geq \mathbf{0}} f(\alpha, \mathcal{I}) \quad (12a)$$

where

$$f(\alpha, \mathcal{I}) = \mathcal{L}(\Phi\alpha, \mathcal{I}) + ur(\alpha) \quad (12b)$$

$$r(\alpha) = \sum_{i=1}^p \sqrt{\sum_{j \in \mathcal{N}_i} (\alpha_i - \alpha_j)^2} + \mathbb{I}_{[0, +\infty)}(\alpha) \quad (12c)$$

are the penalized NLL objective function and the density-map regularization term that enforces nonnegativity and sparsity of the signal α in an the total-variation (TV) domain. Here, $u > 0$ is a scalar tuning constant. The first summand in (12c) is an isotropic TV-domain sparsity term and \mathcal{N}_i is index set of neighbors of the i th element of α , where elements of α are arranged to form a 2D image [21]. In [8], we have used the sparsity regularization term in the form of the ℓ_1 -norm of discrete wavelet transform (DWT) coefficients of α .

Minimization Algorithm

The parameters that we wish to estimate are naturally divided into two blocks, α and \mathcal{I} . The large size of α prohibits effective second-order methods under the TV regularization, whereas \mathcal{I} has much smaller size and only nonnegative constraints, thus allowing for more sophisticated solvers, such as the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) approach [22, Sec. 4.3.3.4] that we adopt here. In addition, the scaling difference between α and \mathcal{I} can be significant so that the joint gradient method for α and \mathcal{I} together would converge slowly. Therefore, we adopt a block coordinate-descent algorithm to minimize $f(\alpha, \mathcal{I})$ in (12b), where the Nesterov's proximal-gradient (NPG) [8] and limited-memory Broyden-Fletcher-Goldfarb-Shanno with box constraints (L-BFGS-B) [23] methods are employed to update estimates of the density-map and mass-attenuation spectrum parameters, respectively.

We minimize the objective function (12b) by alternatively updating α and \mathcal{I} using Step 1) and Step 2), respectively, where Iteration i proceeds as follows:

Step 1) (NPG) Set the mass-attenuation spectrum $\iota(\kappa) = \mathbf{b}(\kappa)\mathcal{I}^{(i-1)}$, treat it as known, and descend the regularized NLL function $f(\alpha, \mathcal{I}^{(i-1)}) = \mathcal{L}_\iota(\alpha) + ur(\alpha)$ [see (10a)] by applying an *NPG step* for α , which yields $\alpha^{(i)}$:

$$\theta^{(i)} = \frac{1}{2} \left[1 + \sqrt{1 + 4(\theta^{(i-1)})^2} \right] \quad (13a)$$

$$\bar{\alpha}^{(i)} = \alpha^{(i-1)} + \frac{\theta^{(i-1)} - 1}{\theta^{(i)}} (\alpha^{(i-1)} - \alpha^{(i-2)}) \quad (13b)$$

$$\alpha^{(i)} = \arg \min_{\alpha} \frac{1}{2\beta^{(i)}} \|\alpha - \bar{\alpha}^{(i)} + \beta^{(i)} \nabla \mathcal{L}_\iota(\bar{\alpha}^{(i)})\|_2^2 + ur(\alpha) \quad (13c)$$

where the minimization (13c) is computed using an inner iteration that employs the TV-based denoising method in [21, Sec. IV] and $\beta^{(i)} > 0$ is an adaptive step size chosen to satisfy the majorization condition:

$$\mathcal{L}_\iota(\alpha^{(i)}) \leq \mathcal{L}_\iota(\bar{\alpha}^{(i)}) + (\alpha^{(i)} - \bar{\alpha}^{(i)})^T \nabla \mathcal{L}_\iota(\bar{\alpha}^{(i)}) + \frac{1}{2\beta^{(i)}} \|\alpha^{(i)} - \bar{\alpha}^{(i)}\|_2^2 \quad (13d)$$

using the following adaptation scheme [8]:

a) if there has been no step size reductions for \mathfrak{n} consecutive iterations, i.e., $\beta^{(i-1)} = \beta^{(i-2)} = \dots = \beta^{(i-\mathfrak{n}-1)}$, start with a larger step size $\beta^{(i)} = \beta^{(i-1)}/\xi$, where $\xi \in (0, 1)$ is a *step-size adaptation parameter*; otherwise start with $\beta^{(i)} = \beta^{(i-1)}$;

b) *backtrack* using the same scaling constant ξ , with goal to find the largest $\beta^{(i)}$ that satisfies (13d).

We restore the monotonicity and improve convergence of NPG steps by applying the “function restart” [24].

Step 2) (BFGS) Set the design matrix $A = \mathbf{b}_\circ^T(\Phi \alpha^{(i)})$, treat it as known, and minimize the regularized NLL function $f(\alpha^{(i)}, \mathcal{I})$ with respect to \mathcal{I} , i.e., solve [see (10b)]

$$\mathcal{I}^{(i)} = \arg \min_{\mathcal{I} \geq \mathbf{0}} \mathcal{L}_A(\mathcal{I}) \quad (14)$$

using the inner L-BFGS-B iteration, initialized by $\mathcal{I}^{(i-1)}$.

We refer to this iteration as the *NPG-BFGS algorithm*. If we *do not* apply the Nesterov's acceleration (13a)–(13b) and use only the proximal-gradient (PG) step (13c) to update the density-map iterates α , i.e., $\bar{\alpha}^{(i)} = \alpha^{(i-1)}$, then the corresponding iteration is the *PG-BFGS algorithm*, see also the numerical examples and Fig. 4b.

Convergence criteria. Define the measures of change of the density map α and the NLL

$$\delta^{(i)} = \|\alpha^{(i)} - \alpha^{(i-1)}\|_2 \quad \delta_{\mathcal{L}}^{(i)} = \left| \mathcal{L}(\alpha^{(i)}, \mathcal{I}^{(i-1)}) - \mathcal{L}(\alpha^{(i-1)}, \mathcal{I}^{(i-1)}) \right| \quad (15a)$$

upon completion of Step 1) in Iteration i . We run the outer iteration between Step 1) and Step 2) until the relative distance of consecutive iterates of the density map α does not change significantly:

$$\delta^{(i)} < \epsilon \|\alpha^{(i)}\|_2 \quad (15b)$$

where $\epsilon > 0$ is the convergence threshold. The convergence criteria for the inner TV-denoising and L-BFGS-B iterations used to solve (13c) and (14) in Iteration i are based on the relative changes $\delta^{(i-1)}$ and $\delta_{\mathcal{L}}^{(i)}$, respectively:

$$\|\alpha^{(i,k)} - \alpha^{(i,k-1)}\|_2 < \eta_{\alpha} \delta^{(i-1)}, \quad \left| \mathcal{L}_A(\mathcal{I}^{(i,k)}) - \mathcal{L}_A(\mathcal{I}^{(i,k-1)}) \right| \leq \eta_{\mathcal{I}} \delta_{\mathcal{L}}^{(i)} \quad (15c)$$

where k are the inner-iteration indices and the convergence tuning constants $\eta_{\alpha} \in (0, 1)$ and $\eta_{\mathcal{I}} \in (0, 1)$ are chosen to trade off the accuracy and speed of the inner iterations and provide sufficiently accurate solutions to (13c) and (14). Here, $\delta^{(i-1)}$, the change of α within Iteration $i - 1$, provides a good guideline for setting the convergence criterion of the inner iteration for α in Step 1) of Iteration i . Since our main goal is to estimate the density map α , the benefit to Step 1) provided by Step 2) through the minimization of $\mathcal{L}_A(\mathcal{I})$ in (14) is more important than the reconstruction of \mathcal{I} itself. Hence, we select $\delta_{\mathcal{I}}^{(i)}$, the change of the NLL within Step 1), as a convergence metric for the inner iteration for \mathcal{I} in Step 2) of Iteration i . Both inner iterations have the maximum number iteration set to n_{sub} . In summary, the outer and inner convergence tuning constants are:

$$(\epsilon, \eta_{\alpha}, \eta_{\mathcal{I}}, n_{\text{sub}}). \quad (15d)$$

Initialization. We initialize the density-map and mass-attenuation coefficient vector iterates as follows [8]:

$$\alpha^{(0)} = \alpha^{(-1)} = \hat{\alpha}_{\text{FBP}}, \quad \theta^{(0)} = 0 \quad (16a)$$

$$\mathcal{I}_j^{(0)} = \begin{cases} \max_n \mathcal{E}_n / b_j^L(0), & j = \lceil 0.5(J+1) \rceil \\ 0, & \text{otherwise} \end{cases} \quad (16b)$$

where $\hat{\alpha}_{\text{FBP}}$ is the standard filtered backprojection (FBP) reconstruction [17, Ch. 3]. We select the initial step size $\beta^{(0)}$ using the Barzilai-Borwein method [25]. Plugging the initialization (16b) into (7) yields $\mathcal{I}^{\text{in}}(\mathcal{I}^{(0)}) = \max_n \mathcal{E}_n$ and the initial estimate $(\alpha^{(0)}, \mathcal{I}^{(0)})$ corresponds approximately to a monochromatic X-ray model; more precisely, it is a polychromatic X-ray model with a narrow mass attenuation spectrum $b_{\lceil 0.5(J+1) \rceil}(\kappa)$. It is also desirable to have the main lobe of the estimated spectrum at the center, which is why the nonzero element of $\mathcal{I}^{(0)}$ is placed in the middle position.

NUMERICAL EXAMPLES

We compare the NPG-BFGS and linear FBP methods by applying them to reconstruct two industrial objects containing defects, labeled C-I and C-II, from real fan-beam projections. The projection matrix Φ used by both methods has been constructed directly on graphics processing unit (GPU) with full circular mask [26] and the multi-thread version on central processing unit (CPU) is also available; see <https://github.com/isucsp/imgRecSrc>.

Figure 2 compares FBP and NPG-BFGS reconstructions from sinograms of C-I and C-II objects. For a comprehensive set of simulated examples comparing our reconstruction approach with existing methods under the lognormal noise model, see [8].

The collected measurements \mathcal{E} have been normalized by division of their maximum such that $\max_n \mathcal{E}_n = 1$, which stabilizes the magnitude of NLL by scaling it with a constant, see (9). We set the B1-spline tuning constants (5d) to satisfy

$$q^J = 10^3, \quad \kappa_{\lceil 0.5(J+1) \rceil} = 1, \quad J = 20 \quad (17a)$$

which ensure sufficient coverage (three orders of magnitude) and resolution (20 basis functions) of the basis-function representation of the mass-attenuation spectrum and centering its support around 1. We set the convergence and adaptive step-size tuning constants for the NPG-BFGS method as

$$(\epsilon, \eta_{\alpha}, \eta_{\mathcal{I}}, n_{\text{sub}}) = (10^{-6}, 10^{-3}, 10^{-2}, 20), \quad (\mathfrak{n}, \xi) = (4, 0.5) \quad (17b)$$

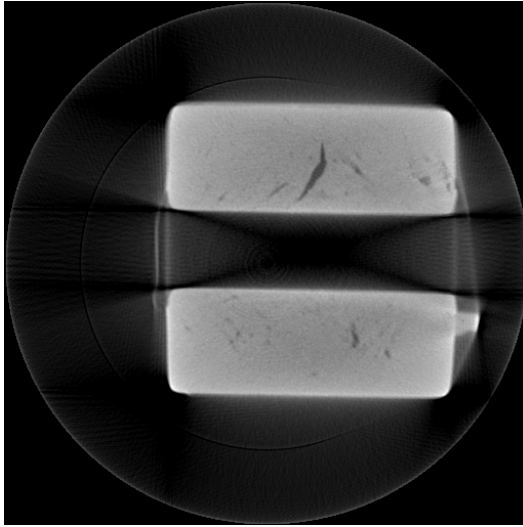
limit the number of outer iterations to 4000 at most, and select the regularization constant

$$u = 10^{-5} \quad (17c)$$

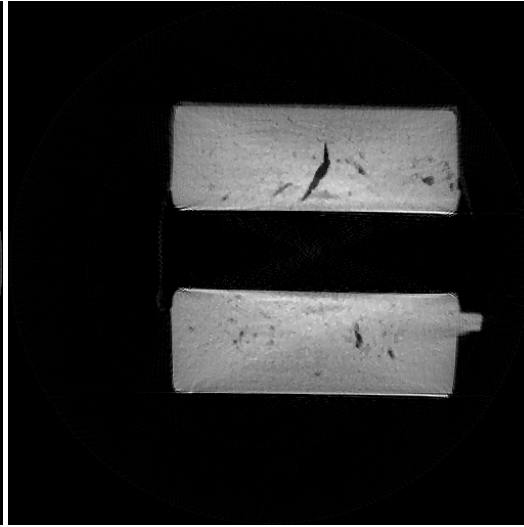
unless specified otherwise. Due to the ambiguity, the NPG-BFGS reconstructions can be determined only up to a scaling constant. Hence, we show the reconstructed NPG-BFGS density maps normalized by their own maximum pixel value.

The C-I data set consists of 360 equi-spaced fan-beam projections with 1° separation collected using an array of 694 detectors, with distance of X-ray source to the rotation center equal to 3492 times the detector size. Figures 2a and 2b show 512×512 density-map image reconstructions of object C-I using the FBP and NPG-BFGS methods, respectively. The linear FBP reconstruction, which does not account for the polychromatic nature of the X-ray source, suffers from severe streaking (shading that occupies the empty area) and cupping (high intensity along the object's border) artifacts whereas the NPG-BFGS reconstruction removes these artifacts thanks to accounting for the polychromatic X-ray source.

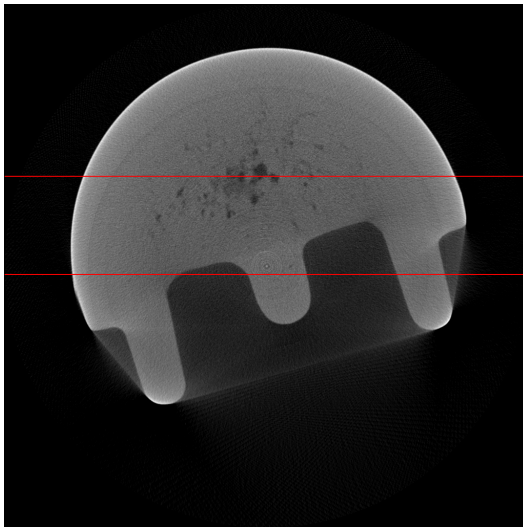
The C-II data set consists of 360 equi-spaced fan-beam projections with 1° separation collected using an array of 1380 detectors, with distance of X-ray source to the rotation center equal to 8696 times the detector size. Figures 2c and 2d show 1024×1024 density-map image reconstructions of object C-II by the FBP and NPG-BFGS methods, respectively. The NPG-BFGS reconstruction removes the streaking and cupping artifacts exhibited by the linear FBP, with enhanced



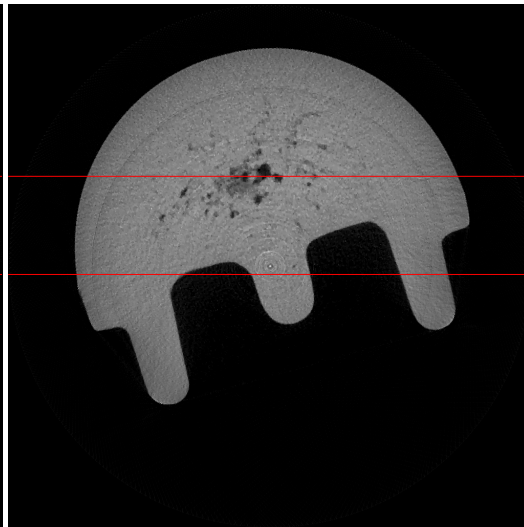
(a) FBP



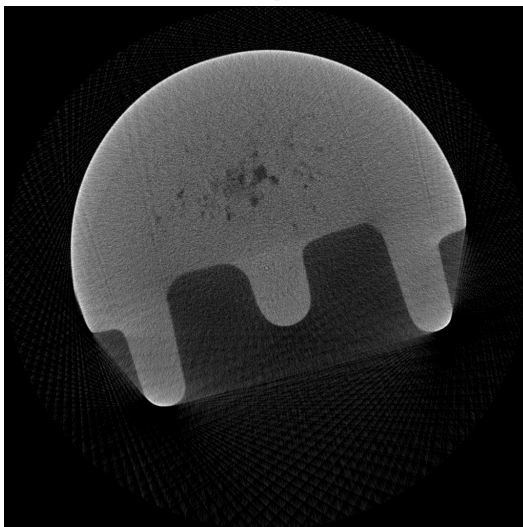
(b) NPG-BFGS



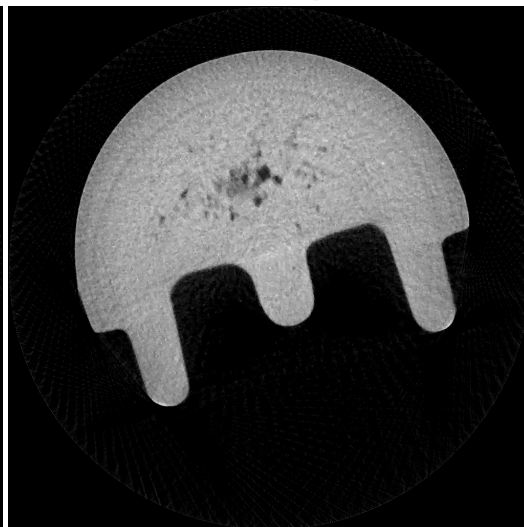
(c) FBP (360 projections)



(d) NPG-BFGS (360 projections)



(e) FBP (120 projections)



(f) NPG-BFGS (120 projections)

FIGURE 2. C-I and C-II object reconstructions from fan-beam projections using the FBP and NPG-BFGS methods.

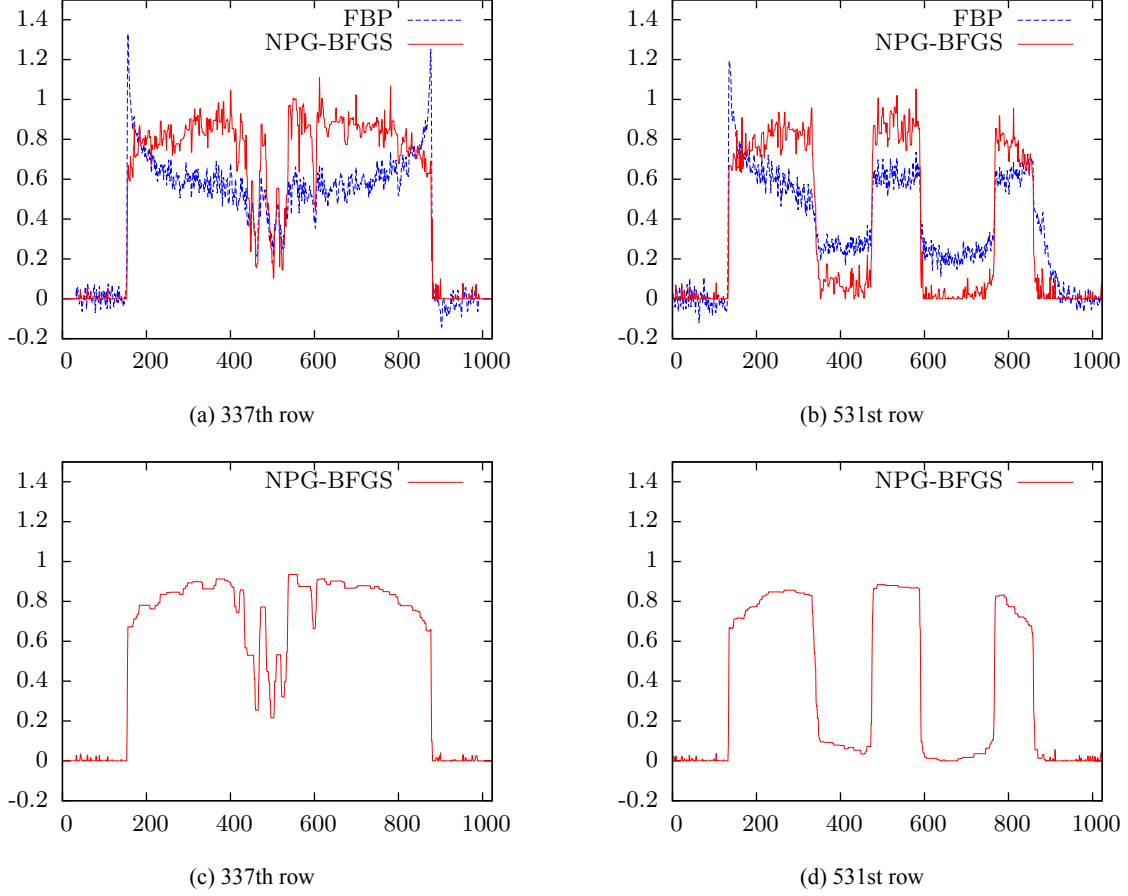


FIGURE 3. C-II object reconstruction profiles with NPG-BFGS reconstructions using (a)–(b) $u = 10^{-5}$ and (c)–(d) $u = 10^{-4}$.

contrast for both the inner defects and object boundary. Figures 2e and 2f show the FBP and NPG-BFGS reconstructions from downsampled C-II data set with 120 equi-spaced fan-beam projections with 3° separation. The FBP reconstruction in Fig. 2e exhibits both beam-hardening and aliasing artifacts. In contrast, the NPG-BFGS reconstruction in Fig. 2f does not exhibit these artifacts because it accounts for the polychromatic X-ray source and employs signal-sparsity regularization in (12c). Indeed, if we reduce the regularization constant u sufficiently, the aliasing effect will occur in the NPG-BFGS reconstruction in Fig. 2f as well.

Figure 3 shows the reconstruction profiles of the 337th and 531th rows highlighted by the red horizontal lines across Figs. 2c and 2d. Noise in the NPG-BFGS reconstructions can be reduced by increasing the regularization parameter u : Figures 3c and 3d show the corresponding NPG-BFGS reconstruction profiles for $u = 10^{-4}$, which is 10 times that in Figs. 3a and 3b.

In Fig. 4a, we show the scatter plots with 1000 randomly selected points representing FBP and NPG-BFGS reconstructions of the C-II object from 360 projections. Denote by $(\hat{\alpha}, \hat{\mathcal{I}})$ the estimate of (α, \mathcal{I}) obtained upon convergence of the NPG-BFGS iteration. The y -coordinates in the scatter plots are the *noisy* measurements in log scale $-\ln \mathcal{E}_n$ and the corresponding x -coordinates are the monochromatic projections $\phi_n^T \hat{\alpha}_{\text{FBP}}$ (red) and $\phi_n^T \hat{\alpha}$ (green) of the estimated density maps. If $(\hat{\alpha}, \hat{\mathcal{I}})$ are the underlying true parameters for the mass-attenuation spectrum model, $-\ln[\mathbf{b}^L(\cdot)\hat{\mathcal{I}}]$ in Fig. 4a is the inverse linearization function that maps monochromatic projections to *noiseless* polychromatic projections $-\ln \mathcal{I}_n^{\text{out}}$. The vertical-direction differences between the NPG-BFGS scatter plot and the corresponding linearization curve show good fit between the measurements and our model. Since FBP assumes linear relation between $-\ln_o \mathcal{I}^{\text{out}}$ and $\Phi \alpha$, its scatter plot (red) can be fitted by a straight line $y = x$, as shown in Fig. 4a. A few points in the FBP scatter plot with $\ln \mathcal{E}_n = 0$ and positive monochromatic projections indicate severe streaking artifacts. Observe relatively large residuals with bias, which remain even if more sophisticated linear models, e.g., iterative algorithms with sparsity and nonnegativity constraints, were adopted, thereby necessitating the need for accounting for the polychromatic source; see the numerical examples in [8].

We now illustrate the advantage of using Nesterov’s acceleration in Step 1) of our iteration. Figure 4b shows the objective

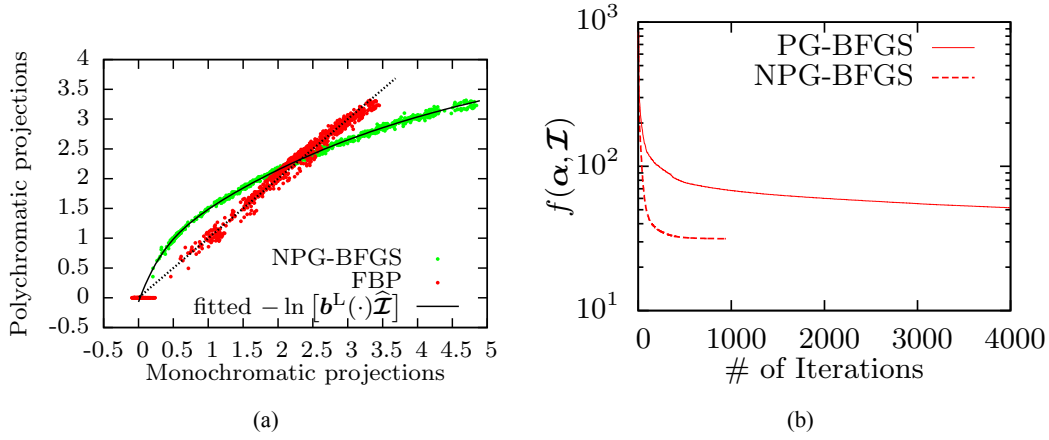


FIGURE 4. (a) Polychromatic measurements as functions of monochromatic projections and corresponding fitted inverse linearization curves, and (b) the objective function as a function of iteration index i .

$f(\alpha, \mathcal{I})$ as a function of outer iteration index i for the NPG-BFGS and PG-BFGS methods applied to C-II reconstruction from 360 projections. Thanks to the Nesterov's acceleration (13b), NPG-BFGS is at least 10 times faster than PG-BFGS, which runs until it reaches the maximum-iteration limit.

CONCLUSION

We developed a polychromatic X-ray CT reconstruction method for Poisson noise that requires no additional information than the conventional FBP method, and yet is capable of correcting beam-hardening *and* aliasing artifacts. We proposed a block coordinate descent iterative algorithm to minimize a penalized NLL, with TV and nonnegativity signal penalty terms. Numerical experiments using real X-ray CT data were presented.

Observe that the NPG-BFGS reconstructions of C-I and C-II have higher contrast around the inner region where cracks reside. Indeed, our reconstructions have slightly higher intensity in the center, which is likely due to the detector saturation that lead to measurement truncation; other possible causes could be scattering or noise model mismatch. (We have replicated this slight non-uniformity by applying our reconstruction to simulated truncated measurements.) This effect is visible in C-I reconstruction in Fig. 2b and is barely visible in the C-II reconstruction in Fig. 2d, but can be observed in the profiles in Fig. 3. We leave further verification of causes and potential correction of this problem to future work and note that this issue does not occur in simulated-data examples that we constructed.

Future work will also include extending our parsimonious mass-attenuation parameterization to multiple materials and developing corresponding reconstruction algorithms.

ACKNOWLEDGMENTS

This work was supported by the NSF Award CCF-1421480 and NSF Industry-University Cooperative Research Program, Center for Nondestructive Evaluation (CNDE), Iowa State University. The authors are grateful to Dr. Joseph N. Gray, CNDE, Iowa State University, for providing real X-ray CT data used in the numerical examples.

REFERENCES

1. B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Trans. Med. Imag.* **20**, 999–1008 (2001).
2. I. A. Elbakri and J. A. Fessler, "Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation," *Phys. Med. Biol.* **48**, 2453–2477 (2003).
3. I. A. Elbakri and J. A. Fessler, "Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography," in *Proc. SPIE Med. Imag.* Vol. 5032 (Feb. 2003), pp. 1839–1850.

4. B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson, "Properties of preprocessed sinogram data in X-ray computed tomography," *Med. Phys.* **33**, 3290–3303 (2006).
5. I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.* **21**, 89–99 (2002).
6. G. Van Gompel, K. Van Slambrouck, M. Defrise, K. Batenburg, J. de Mey, J. Sijbers, and J. Nuyts, "Iterative correction of beam hardening artifacts in CT," *Med. Phys.* **38**, S36–S49 (2011).
7. J. Xu and B. M. Tsui, "Quantifying the importance of the statistical assumption in statistical X-ray CT image reconstruction," *IEEE Trans. Med. Imag.* **33**, 61–73 (2014).
8. R. Gu and A. Dogandžić, "Polychromatic sparse image reconstruction and mass attenuation spectrum estimation via B-spline basis function expansion," in *Review of Progress in Quantitative Nondestructive Evaluation*, eds. D. E. Chimenti and L. J. Bond, (American Institute of Physics 1650, Melville, NY), **34**, 1707–1716 (2015).
9. J. D. Evans, B. R. Whiting, D. G. Polite, J. A. O'Sullivan, P. F. Klahr, and J. F. Williamson, "Experimental implementation of a polyenergetic statistical reconstruction algorithm for a commercial fan-beam CT scanner," *Phys. Medica* **29**, 500–512 (2013).
10. R. Redus, J. Pantazis, T. Pantazis, A. Huber, and B. Cross, "Characterization of CdTe detectors for quantitative X-ray spectroscopy," *IEEE Trans. Nucl. Sci.* **56**, 2524–2532 (2009).
11. D. Zhang, X. Li, and B. Liu, "X-ray spectral measurements for tungsten-anode from 20 to 49 kVp on a digital breast tomosynthesis system," *Med. Phys.* **39**, 3493–3500 (2012).
12. Y. Lin, J. C. Ramirez-Giraldo, D. J. Gauthier, K. Stierstorfer, and E. Samei, "An angle-dependent estimation of CT x-ray spectrum from rotational transmission measurements," *Med. Phys.* **41**, 062104 (2014).
13. R. Gu and A. Dogandžić, "Sparse X-ray CT image reconstruction and blind beam hardening correction via mass attenuation discretization," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.* (Dec. 2013), pp. 244–247.
14. G. M. Lasio, B. R. Whiting, and J. F. Williamson, "Statistical reconstruction for X-ray computed tomography using energy-integrating detectors," *Phys. Med. Biol.* **52**, 2247 (2007).
15. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3rd ed. (McGraw-Hill, New York, 1957).
16. J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography," *Phys. Med. Biol.* **58**, R63–R96 (2013).
17. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE Press, New York, 1988).
18. G. T. Herman, "Correction for beam hardening in computed tomography," *Phys. Med. Biol.* **24**, 81–106 (1979).
19. L. Zanni, A. Benfenati, M. Bertero, and V. Ruggiero, "Numerical methods for parameter estimation in Poisson data inversion," *J. Math. Imaging Vis.* **52**, 397–413 (2015).
20. P. McCullagh and J. Nelder, *Generalized Linear Models*, 2nd ed. (Chapman & Hall, New York, 1989).
21. A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.* **18**, 2419–2434 (2009).
22. R. A. Thisted, *Elements of Statistical Computing* (Chapman & Hall, New York, 1989).
23. R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.* **16**, 1190–1208 (1995).
24. B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.*, 1–18 (2013).
25. J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.* **8**, 141–148 (1988).
26. A. Dogandžić, R. Gu, and K. Qiu, "Mask iterative hard thresholding algorithms for sparse image reconstruction of objects with known contour," in *Proc. Asilomar Conf. Signals, Syst. Comput.* (Nov. 2011), pp. 2111–2116.